Classification Decision Trees vs Regression

Here's a summary of our conversation:

1. **Classification Decision Tree Explanation (Q1)**:
   - We discussed the purpose of classification decision trees, real-world applications, and compared their approach to making predictions with multiple linear regression.
2. **Metrics Analysis (Q2)**:
   - Explained and computed metrics like sensitivity, specificity, and accuracy for classification models, with examples of real-world use cases for each.
3. **Dataset Preprocessing (Q3)**:
   - Addressed cleaning the dataset by removing unnecessary columns, handling missing values, and adjusting data types.
4. **Training a Decision Tree (Q4)**:
   - Created and visualized a decision tree using one feature (`List Price`), performed an 80/20 split, and explained how predictions were made.
5. **Improved Model with More Features (Q5)**:
   - Expanded the decision tree (`clf2`) to include `NumPages`, `Thick`, and `List Price`, increased `max_depth` to 4, and interpreted its visualization.
6. **Confusion Matrices and Metrics (Q6)**:
   - Generated confusion matrices for `clf` and `clf2`, calculated sensitivity, specificity, and accuracy, and compared the models' performances.
7. **Training Data Confusion Matrices (Q7)**:
   - Discussed why `clf2` outperformed `clf` on training data due to its additional features, making it better at capturing relationships.
8. **Feature Importances (Q8)**:
   - Visualized and reported the feature importances for `clf2`, identifying that **List Price** was the most influential predictor.
9. **Comparison of Interpretations (Q9)**:
   - Compared the interpretation of coefficients in linear regression with feature importances in decision trees, focusing on their differences in explaining variable influence.

Let me know if you'd like to revisit any topic!